NONPARAMETRIC TESTS FOR LONGITUDINAL DATA

by

LEI DONG

B.S., Iowa State University, 2005

A REPORT

submitted in partial fulfillment of the requirements for the degree

MASTER OF SCIENCE

Department of Statistics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2009

Approved by:

Major Professor
Dr. Haiyan Wang

Copyright

Abstract

The purpose of this report is to numerically compare several tests that are applicable to longitudinal data when the experiment contains a large number of treatments or experimental conditions. Such data are increasingly common as technology advances. Of interest is to evaluate if there is any significant main effect of treatment or time, and their interactions. Traditional methods such as linear mixed-effects models (LME), generalized estimating equations (GEE), Wilks' lambda, Hotelling-Lawley, and Pillai's multivariate tests were developed under either parametric distributional assumptions or the assumption of large number of replications. A few recent tests, such as Zhang (2008), Bathke & Harrar (2008), and Bathke & Harrar (2008) were specially developed for the setting of large number of treatments with possibly small replications. In this report, I will present some numerical studies regarding these tests. Performance of these tests will be presented for data generated from several distributions.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I would like to thank my major advisor, Dr. Haiyan Wang. Without her close guidance, invaluable assistance and advice, and tremendous patience, I cannot reach this point. I am indebted to her in my whole career development.

I would like to thank Dr. Weixing Song and Dr. Gary Gadbury for serving on my committee. Their consistent supports and their efforts for reviewing and commenting on my research work are highly appreciated.

I would like to thank all faculty members in Department of Statistics at Kansas State University. I learned a great deal from taking their courses. Their kindness and willingness to help me will never be forgotten.

I also want to say thank Dr. Ruth Welti for her supports. I worked in Kansas Lipidomics Research Center two years and a half as a research assistant. During working in the lab, I learned lots of most useful knowledge of applied statistic**.**

Finally, I would like to thank my wife, Yu Jiang. Her trust and consistent support have made it possible to complete the research work.

CHAPTER 1 - Introduction

With the development of technologies, “omics” studies, such as genomics, proteomics and lipidomics, are becoming more and more popular in biological studies. In these studies, the number of variables observed over time, such as gene, protein, and lipid species, are very large, but the sample size per treatment is small. For example, in a typical microarray study, the number of genes measured is usually ten thousand. Due to the high cost of measurement and other reasons, the number of replicates at each time point is often three or four. This type of data is called the high dimensional longitudinal data.

Due to the large number of variables involved, it is often unrealistic to assume common variance/covariance structure for all the variables throughout all experimental conditions. For this reason, traditional methods such as linear mixed effects model may not be appropriate for analyzing high dimensional longitudinal data. With the small replications, some classical methods such as generalized estimating equations (GEE) approach tend to be invalid or have low power in current setting. GEE was proposed in Liang and Zerger (1986) to extend quasi-likelihood for exponential family by Wedderburn (1974) for analyzing longitudinal data. Subsequently, Liang and Zeger (1994) had given a description of several different methods to analyze the longitudinal data. They developed the models for analysis of longitudinal data from different distributions of response variables. For example, for normally distributed variables, they recommended to use the general linear models; but for binary and count data, they recommended to use the generalized linear models (GLMs). Kshirsagar and Smith (1995) had a study about use of growth curves to analyze longitudinal data. Also, Verbeke and Molenberghs (2000) provided a comprehensive treatment of linear mixed models for continuous longitudinal data. All of these studies are based on parametric model within the exponential family.

Harrar & Bathke(2008), and Bathke & Harrar(2008) did some studies about nonparametric methods in multivariate factorial designs with a large number of treatments. They proposed tests that were non-parametric analogues of the ANOVA-type, Bartlett-Nanda-Pillai's, and Hotelling-Lawley statistics. The simulations examined in their study indicated that, none of the three tests they considered are uniformly best.

Zhang (2008) also provided a set of nonparametric tests to evaluate the effect of treatments, time, and their interactions when there are a large number of heteroscedastic

treatment levels. All three references mentioned in this paragraph derived their asymptotic theory under the non-classical setting in which the number of treatments is large while the sample sizes are small.

In this report, we will first briefly review the few nonparametric methods in aforementioned references. Then we will present the simulation study to investigate their numerical performance compared to traditional parametric methods in terms of type I error and power estimate. GEE, LME, Wilks' lambda, Pillai's, and Hotelling-Lawley multivariate tests, as well as their nonparametric analogs by Harrar & Bathke(2008), and Bathke & Harrar(2008), and Zhang's nonparametric tests will be considered in the comparison. We will examine the robustness of type I error of these methods for data generated from a few distributions including normal, gamma and Possion. The power estimates are compared for mixture distribution with normal and gamma components.

CHAPTER 2 - Background Review

2.1 Introduction of longitudinal data set

A longitudinal study is defined as a study in which the response for each experimental unit in the study is observed on two or more occasions. A longitudinal data set contains repeated observations on each experimental unit. People are usually interested in the pattern of change over time or the dependence of the outcome on the covariates.

Denote the response from the k^{th} subject in the i^{th} treatment at the j^{th} time point as X_{ijk}. The observations from the same subject can be denoted as a random vector $X_{ik}=(X_{i1k} , \ldots, X_{ibk})'$, i=1,2,…,a, and k=1, …,n_i. All subjects are typically assumed to be independent. The data can be stored in two formats in computer, the long format (Table 2.1) in which each row gives a single observation from a subject, and the wide format (Table 2.2) in which all observations from the same subject are stored in the same row.

Treatments	Times	Subjects	Responses
i=1	1	1	x_{111}
		2	x_{112}
		…	…
		n_1	x_{11n1}
	…	…	…
	b	1	x_{1b1}
		2	x_{1b2}
		…	…
		n_1	x_{1bn1}
…	…	…	…
i=a	1	1	x_{a11}
		2	x_{a12}
		…	…
		n_a	x_{a1na}
	…	…	…
	b	1	x_{ab1}
		2	x_{ab2}
		…	…
		n_a	x_{abna}

Table 2-1 Long format

Treatments	Subjects	Time			
		1	2	…	b
i=1	1	x_{111}	x_{121}	…	x_{1bt1}
	2	x_{112}	x_{122}	…	x_{1b2}
	…	…	…	…	…
	n_1	x_{11n1}	x_{12n1}	…	x_{1bn1}
…	…	…	…	…	…
i=a	1	x_{111}	x_{121}	…	x_{ab1}
	2	x_{112}	x_{122}	…	x_{ab2}
	…	…	…	…	…
	n_a	x_{11na}	x_{12na}	…	x_{abna}

Table 2-2 Wide format

2.2 Model Specification

We have longitudinal data with some discrete factors, such as genotypes, abiotic stresses. For ease of notation, we use a single factor treatment to denote a composite factor whose levels are level combinations of all discrete factors. In this report, we use factors and variables interchangeably. The number of factor levels is large and the number of time points is fixed. In high dimensional biological study, the data could be unbalanced. We follow the typical experimental setting of small number of subjects. Assume the number of subjects in treatment i is n_i. For the k^{th} subject, the datum at j^{th} time point of i^{th} factor level is denoted by $X_{ijk}, i = 1,2,...,a; j = 1,2,...,b; k = 1,2,...,n_i$. We are interested in testing if there is any significant differences among the treatments or time points, and whether there exits treatment effects over time, i.e. to test the interaction effect between treatments and time points.

Following the notation of Wang & Akritas (2009) and Wang et al. (2009), a nature model is

$$X_{ijk} = \mu + \alpha_i + \beta_j + B_{ik} + \gamma_{ij} + e_{ijk},$$

where μ is the overall mean, α_i is the effect of the i^{th} treatment, β_j is the effect of j^{th} time point, γ_{ij} is the interaction effect of i^{th} treatment at j^{th} time point. The model parameters are defined through decomposition of the observation from a randomly selected subject $S_{k(i)}$ in the i^{th} treatment:

$$\mu_{ijS_{k(i)}} = E(X_{ijS_{k(i)}}|S_{k(i)}),\quad \mu = (ab)^{-1}\sum_{i=1}^{a}\sum_{j=1}^{b}E(\mu_{ijS_{k(i)}}),\quad \alpha_i = b^{-1}\sum_{j=1}^{b}E(\mu_{ijS_{k(i)}})-\mu,$$

$$B_{ik} = b^{-1}\sum_{j=1}^{b}\mu_{ijS_{k(i)}}-\mu-\alpha_i,\quad \beta_j = a^{-1}\sum_{i=1}^{a}E(\mu_{ijS_{(i)}})-\mu,\quad \gamma_{ij} = E(\mu_{ijS_{k(i)}})-\mu-\alpha_i-\beta_j,$$

with the following constraints

$$\sum_i \alpha_i = \sum_j \beta_j = \sum_i \gamma_{ij} = \sum_j \gamma_{ij} = E(B_{ik}) = E(e_{ijk}) = 0.$$

The term e_{ijk} includes the measurement error and some random subject by time interactions not explicitly written in above model

$$D_{ijS_{k(i)}} = \mu_{ijS_{k(i)}}-\mu-\alpha_i-\beta_j-B_{ik}-\gamma_{ij}.$$

A justification of the hidden random interactions, as pointed out by Wang et al. (2009) is from the Bayes regression model in which both the data and the model parameters come from some unknown stochastic process (Morris 1983), or from a multi-level mixed model (hierarchical linear model) used in educational research where the response is related to a set of subject level predictors via a linear model for each subject at the first stage, and the model parameters from the first stage are used as the response variables in a second stage model (Raudenbush and Bryk (2002)). The subject specific random effect B_{ik} and the error e_{ijk} are not independent due to the random interactions included in e_{ijk} even though they are typically assumed to be independent in models of most text books.

As our interest only lies in the fixed effects, we can combine the random effect B_{ik} and the error e_{ijk} term together into a single composite error term ε_{ijk}. That is, a general marginal model without distributional assumption can be written as

$$X_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk} \qquad (2.2.1)$$

where ε_{ijk} is the error term with mean 0 and $(\varepsilon_{i1k,}\cdots\cdots,\varepsilon_{ibk})$' has unknown covariance matrix following the fact that observations from the same subject are correlated. The error terms are not

necessary following a normal distribution, and the model allows heteroscedastic treatment effects. This is necessary particularly due to a large number of factor levels.

2.3 Multivariate and Nonparametric Tests in Longitudinal Data Analysis

In this subsection, we review the three classical multivariate tests, i.e., Wilk's Lambda, Bartlett-Nanda-Pillai's Trace, and Hotelling-Lawley's Trace, and their nonparametric analogs from Harrar & Bathke(2008), and Bathke & Harrar(2008) for analyzing longitudinal data. The difference is mainly in calculation of the degrees of freedom.

2.3.1 The Multivariate Tests

For multivariate tests, the raw data are measurements from multiple variables that may be correlated. The structure of multivariate data can be expressed as in Table 2.3:

Treatment 1				*Treatment* 2				...	*Treatment* *a*			
x_{11}^1	x_{12}^1	...	$x_{1n_1}^1$	x_{21}^1	x_{22}^1	...	$x_{2n_2}^1$	...	x_{a1}^1	x_{a2}^1	...	$x_{an_a}^1$
x_{11}^2	x_{12}^2	...	$x_{1n_1}^2$	x_{21}^2	x_{22}^2	...	$x_{2n_2}^2$	...	x_{a1}^2	x_{a2}^2	...	$x_{an_a}^2$
...	...	...	...	...	...	...	...	...	...	...	...	...
x_{11}^p	x_{12}^p	...	$x_{1n_1}^p$	x_{21}^p	x_{22}^p	...	$x_{2n_2}^p$	...	x_{a1}^p	x_{a2}^p	...	$x_{an_a}^p$

Table 2-3 The data format for multivariate tests

where, p is the number of different variables (the time points), a is the number of different conditions (the treatment), and n_i is the number of subjects per condition. To apply the multivariate tests to longitudinal data, we can think of the repeated measurements from the same subjects as measurements from the p variables that may be correlated. For convenience, we use vector notation $\mathbf{X}_{ik}=(X_{ik}^1, \ldots., X_{ik}^p)'$ and denote

$\mathbf{X}_{i.}$= total of all vectors from the ith treatment condition

$\mathbf{X}_{..}$ = overall total

$\overline{X}_{i.} = X_i / n$ mean vector of the i[th] treatment condition

$\overline{X}_{..} = X_{..} / n$ overall mean vector

p = Number of dependent variables

a = Number of treatments

q = rank (C (X_{ik}´ X_{ik})$^{-1}$C´) = a-1

The multivariate tests can only test for the hypothesis regarding treatment effects

$$H_0(A): all \quad \alpha_i = 0, \quad for \quad i = 1,2,...,a.$$

The test of no main time or interaction effects can not be conducted through multivariate tests.

2.3.1.1 The Classical Multivariate Tests:

a. Wilks' Lambda

The Wilks' Lambda test statistic is define as

$$T_{WL} = \Lambda = \frac{\det(E)}{\det(E+H)} = \prod \frac{1}{1+\gamma_i}$$

where

$$E = n\sum_{i=1}^{a}(\overline{X}_{i.} - \overline{X}_{..})(\overline{X}_{i.} - \overline{X}_{..})' = \sum_{i=1}^{a}\frac{1}{n}X_{i.}X_{i.}' - \frac{1}{na}X_{..}\overline{X}_{..}$$

$$H = \sum_{i=1}^{a}\sum_{j=1}^{b}(X_{ij} - \overline{X}_{i.})(X_{ij} - \overline{X}_{i.})' = \sum_{ij}X_{ij}X_{ij}' - \sum_{i}X_{i.}X_{i.}'$$

and $\gamma_1 > \gamma_2$ ... are the ordered eigenvalues of E^{-1}, this can be transformed to an approximate F-statistic:

$$F_W = \frac{\left(1-\Lambda^{1/t}\right)}{\Lambda^{1/t}}\frac{rt-2u}{pq}$$ with degrees of freedom: pq and rt-2u,

Where,

p = Number of dependent variables

k = Number of treatments

q = rank (C (X_{ik}´ X_{ik})$^{-1}$C´) = a-1

r = (N-k) – (p+k+1)/2

u = (pq-2)/4

$$t = \begin{cases} \sqrt{(p^2q^2-4)/(p^2+q^2-5)} & if \quad p^2+q^2-5>0 \\ 1 & otherwise \end{cases}$$

b. Bartlett-Nanda-Pillai's Trace

The Bartlett-Nanda-Pillai's trace statistic is define as V = trace (H $(H+E)^{-1}$) = $\Pi \frac{\gamma_i}{1+\gamma_i}$,

where $(H+E)^{-1}$ is the Moore-Penrose generalized inverse of (H+E), E and H are same as in Wilks' Lambda statistic. This can be transformed to an approximate F-statistic as

$F_P = \left(\frac{2n+s+1}{2m+s+1}\right)\left(\frac{V}{s-V}\right)$ with degrees of freedom s (2m+s+1) and s (2n+s+1),

Where,

n = $(N-a-p^{-1})/2$

m = $(|p-q|^{-1})/2$

s = min (p,q)

c. Hotelling-Lawley's Trace

The Hotelling-Lawley's trace statistic is defined as U = trace $(HE^{-1}) = \Pi\gamma_i$, where, $(HE)^{-1}$ is the Moore-Penrose generalized inverse of HE. This can be transformed to an approximate F-statistic: $F_{HL} = \frac{2(sn+1)U}{s^2(2m+s+1)}$ with degrees of freedom: s (2m+s+1) and 2(sn + 1).

2.3.1.2 Nonparametric Analogs of Multivariate Tests (Harrar & Bathke(2008), Bathke & Harrar(2008))

Harrar & Bathke(2008) and Bathke & Harrar(2008) both considered rank test statistics based on separate rankings for the p different variables, which was called time points in this report. Bathke & Harrar(2008) considered balanced case while Harrar & Bathke(2008) considered unbalanced case.

For the j^{th} sample in the i^{th} treatment condition, the observation for the k^{th} variable $X^{(k)}_{ij} \sim F^{(k)}_i$, k = 1, … , p, where $F^{(k)}_i$ is the average of left and right continuous version of the cumulative distribution function, $F^{(k)}_i(x) = \frac{1}{2} P(X^{(k)}_{ij} \le x) + \frac{1}{2} (X^{(k)}_{ij} < x)$.

Let $R^{(k)}_{ij}$ be the (mid-) rank of $X^{(k)}_{ij}$ for all N=a n observations $X^{(k)}_{11, \ldots,} X^{(k)}_{an}$ under balanced design (Bathke & Harrar, 2008), or for all $\sum_{i=1}^{a} n_i$ observations $X^{(k)}_{11, \ldots,} X^{(k)}_{ana}$ for the k^{th} variable (Harrar & Bathke, 2008). Denote $R_{ij} = (R^{(1)}_{ij}, \ldots, R^{(p)}_{ij})'$ and $R = (R_{11, \ldots,} R_{1n}, R_{21, \ldots,} R_{an})$, the p x N matrix of rank transforms of all observations based on ranking within each variable. For example, in the one-way layout below, each row is ranked separately.

Treatment 1				*Treatment* 2				...	*Treatment* *a*			
$R^{(1)}_{11}$	$R^{(1)}_{12}$	...	$R^{(1)}_{1n_1}$	$R^{(1)}_{21}$	$R^{(1)}_{22}$	...	$R^{(1)}_{2n_2}$	...	$R^{(1)}_{a1}$	$R^{(1)}_{a2}$	...	$R^{(1)}_{an_a}$
$R^{(2)}_{11}$	$R^{(2)}_{12}$	...	$R^{(2)}_{1n_1}$	$R^{(2)}_{21}$	$R^{(2)}_{22}$	...	$R^{(2)}_{2n_2}$	...	$R^{(2)}_{a1}$	$R^{(2)}_{a2}$	...	$R^{(2)}_{an_a}$
...	...	...	...	...	...	...	...	...	...	...	...	...
$R^{(p)}_{11}$	$R^{(p)}_{12}$	...	$R^{(p)}_{1n_1}$	$R^{(p)}_{21}$	$R^{(p)}_{22}$	...	$R^{(p)}_{2n_2}$	...	$R^{(p)}_{a1}$	$R^{(p)}_{a2}$	...	$R^{(p)}_{an_a}$

Table 2-4 The mid-rank of data ranked separately within each variable

where p is the number of different variables (the time points), a is the number of different conditions (the treatment), and n_i is the number of subjects per condition.

In both Harrar & Bathke(2008) and Bathke & Harrar(2008), the nonparametric hypotheses were stated either in the terms of multivariate distribution or in the marginal distributions. Example, in the nonparametric one-way layout, the multivariate null hypothesis was as H_0: all of F_i are equal, the marginal null hypothesis was as $\hat{H}_0$: all of $F^{(k)}_i$ are equal, where i =1, .., a, k = 1, …, p.

The test statistics H (the mean squares due to treatment (hypothesis mean sum of squares)) and E (the mean squares due to error) based on the quadratic forms as

$$H(R) = \frac{1}{a-1} \sum_{i=1}^{a} n(\overline{R}_{i\cdot} - \overline{R}_{\cdot\cdot})(\overline{R}_{i\cdot} - \overline{R}_{\cdot\cdot})'$$

$$E(R) = \frac{1}{N-a} \sum_{i=1}^{a} \sum_{j=1}^{n} (R_{ij} - \overline{R}_{i\cdot})(R_{ij} - \overline{R}_{i\cdot})'$$

where $\overline{R}_{i\cdot}$ and $\overline{R}_{\cdot\cdot}$ are p x 1 vectors as same as $\overline{R}_{ij}$.

Under null hypothesis H_0: no treatment effect, the standardized multivariate nonparametric test statistics are asymptotically a standard normal with a $\to\infty$, n and p fixed (Bathke & Harrar, 2008).

a. Nonparametric Analog of Hotelling-Lawley's Trace (Harrar & Bathke(2008))

The Hotelling-Lawley type statistic was denoted as $U = tr[(a-1)H((N-a)G)^{-1}]$, then, the Hotelling-Lawley test is

$$HL_a = \sqrt{\frac{a(n-1)}{2np_1}}[tr(HE^{-}) - r_1]$$

where, r_1=rank(E). Here, the approximate Hotelling-Lawley's statistic is the adjusted test statistic $F_{K,D}$ times g, $U = F_{K,D} \times g$, where $g = \frac{p(a-1)(D-2)}{(N-a-p-1)D}$, $K = p(a-1)$, $D = 4 + \frac{p(a-1)+2}{B-1}$, and

$$B = \frac{(N-p-2)(N-a-1)}{(N-a-p)(N-a-p-3)}$$

b. Nonparametric Analog of Bartlett-Nanda-Pillai's Trace (Harrar & Bathke(2008))

The Bartlett-Nanda-Pillai type statistic was denoted as

$$V = tr\{(a-1)H[(a-1)H + (N-a)E]^{-1}\},$$

then, the Bartlett-Nanda-Pillai test is $BNP_a = \sqrt{\frac{a(n-1)}{2nr_1}}(\frac{N-1}{N-a})[(N-1)V - r_2]$ where r_2 is the rank of [(a-1) H+ (N-a) E]. And this can be transformed to an approximate F-statistic as

$F_{BNP\,a} = \frac{(V/\gamma)/v_1}{(1-V/\gamma)/v_2}$ with degrees of freedoms v_1 and v_2 such as

$$v_1 = \frac{p(a-1)}{\gamma(N-1)}[\frac{\gamma(N-a+\gamma-p)(N+2)(N-1)}{(N-a)(N-p)} - 2]$$

$$v_2 = \frac{p-a+\gamma-p}{N}[\frac{\gamma(N-a+\gamma-p)(N+2)(N-1)}{(N-a)(N-p)} - 2]$$

and, γ =min(a-1, p).

2.3.2 Tests of Zhang (2008)

The results of Zhang (2008) were given under model (2.2.1). The following notations will be used in this subsection of the report,

$$\widetilde{X}_{i..} = b^{-1}\sum_{j=1}^{b}\overline{X}_{ij.}, \qquad \widetilde{X}_{.j.} = a^{-1}\sum_{i=1}^{a}\overline{X}_{ij.}$$

$$\sigma_{i,jj1} = Cov(X_{ijk}, X_{ij_1k}) \quad fork, \quad note \quad \sigma_{i,jj} = \mathrm{var}(X_{ijk}) = \sigma^2{}_{ij}$$

$$\sigma_{i,jj_1,j_2j_3} = Cov(X_{ijk}X_{ij_1k}, X_{ij_2k}X_{ij_3k}), \quad note \quad \sigma_{i,jj_1,jj_1} = \sigma^2{}_{i,jj_1}$$

a. Testing statistics

1) For the null hypothesis of no treatment effect

$$H_0(A): all \quad \alpha_i = 0, \quad for \quad i = 1,2,...,a.$$

Zhang (2008) gave a modified F statistic used in mixed ANOVA model

$$F_X(A) = \frac{MST_A}{MSE_A} \tag{I}$$

where

$$MST_A = \frac{1}{a-1}\sum_{i=1}^{a}\sum_{j=1}^{b}(\widetilde{X}_{i..} - \widetilde{X}_{...})^2$$

$$MSE_A = \frac{1}{ab}\sum_{i=1}^{a}\sum_{j,j_1=1}^{b}\frac{1}{n_i(n_i-1)}\sum_{k=1}^{n_i}(X_{ijk} - \overline{X}_{ij.})(X_{ij_1k} - \overline{X}_{ij_1.})$$

2) For the null hypothesis of no time effect

$$H_0(B): all \quad \beta_j = 0, \quad for \quad j = 1,2,...,b.$$

Zhang (2008) considered a more general hypothesis on the contrast effects as

$H_0(B_G): L\beta = 0, \quad where \quad \beta = (\beta_1,...,\beta_b)'$. A modified Wald-type test statistic is used for testing $H_0(B_G)$:

$$W_B = D_B'L'(L\hat{V}_BL')^{-1}LD_B \tag{II}$$

where, $D_B = (\widetilde{X}_{.1.},...,\widetilde{X}_{.b.})'$, and $\hat{V}_B$ is the estimated b*b covariance matrix for D_B, with the value at the j[th] row and j'[th] column be $\hat{V}^{jj'}{}_B$ as

$$\hat{V}^{jj'}{}_B = \frac{1}{a^2}\sum_{i=1}^{a}\frac{1}{n_i(n_i-1)}\sum_{k=1}^{n_i}(X_{ijk}-\overline{X}_{ij.})(X_{ij'k}-\overline{X}_{ij'.}).$$

3) For the null hypothesis of no interaction effects between treatment and time points

$$H_0(AB): all \quad \gamma_{ij}=0, \quad for \quad i=1,2,...,a, \quad and \quad j=1,2,...,b$$

a modified F statistic is given in Zhang (2008)

$$F_X(AB) = \frac{MST_{AB}}{MSE_{AB}} \tag{III}$$

Where,

$$MST_{AB} = \frac{1}{(a-1)(b-1)}\sum_{i=1}^{a}\sum_{j=1}^{b}(\widetilde{X}_{ij.}-\widetilde{X}_{i..}-\widetilde{X}_{.j.}+\widetilde{X}_{...})^2$$

$$MSE_{AB} = \frac{1}{a(b-1)}\sum_{i=1}^{a}\sum_{j=1}^{b}\frac{1}{n_i(n_i-1)}\sum_{k=1}^{n_i}(X_{ijk}-\overline{X}_{ij.})^2 -$$

$$\frac{1}{ab(b-1)}\sum_{i=1}^{a}\sum_{j,j_1=1}^{b}\frac{1}{n_i(n_i-1)}\sum_{k=1}^{n_i}(X_{ijk}-\overline{X}_{ij.})(X_{ij_1k}-\overline{X}_{ij_1.})$$

b. Asymptotic distribution of Zhang (2008)'s test statistics

We state the asymptotic distribution of the test statistics in Zhang (2008).

For testing $H_0(A): all \quad \alpha_i=0, \quad for \quad i=1,2,...,a.$, let F_x (A) be the statistic given in (I). If X_{ijk} has finite fourth moment, then under H_0(A),

$$\frac{\sqrt{a}(F_X(A)-1)}{V_A} \xrightarrow{d} N(0,\ 1) \quad as \quad K\to\infty.$$

Where V_A is the variance component, it is defined as

$$V_A = \sqrt{\tau_A}\Big/\sigma_A, \quad where \quad \tau_A = \frac{1}{ab^2}\sum_{i=1}^{a}\frac{2}{n_i(n_i-1)}\sum_{j,j1,j2,j3}^{b}\sigma_{i,jj_1}\sigma_{i,j_2j_3}, \quad \sigma_A = \frac{1}{ab}\sum_{i=1}^{a}\sum_{j,j1}^{b}\frac{\sigma_{i,jj_1}}{n_i}$$

For testing $H_0(B_G): L\beta = 0,$ *where* L *is a* $J \times p$ *contrst matrix,* $\beta = (\beta_1,...,\beta_b)'$, and 0 is a p dimensional zero vector, let W_B be the statistic given in (II). If X_{ijk} has finite second and fourth moments, then under $H_0(B_G)$,

$W_B \xrightarrow{d} \chi^2_p$ holds for all $n_i \geq 2,\ i = 1,...,\ a.$

For testing $H_0(AB): all\ \ \gamma_{ij} = 0$, let F_x (AB) be the statistic given in (III). If X_{ijk} has finite fourth moment, then under $H_0(AB)$,

$\frac{\sqrt{a}(F_X(AB)-1)}{V_{AB}} \xrightarrow{d} N(0,\ 1),$ holds for $n_i \geq 4$.

where $V_{AB} = \sqrt{\tau_{AB}}/\sigma_{AB}$,

$$\tau_{AB} = \frac{2}{I(J-1)^2}\sum_i^I \frac{1}{n_i(n_i-1)}\left[\sum_{j,j_1}^J \sigma^2_{i,jj_1} + \frac{1}{J^2}\left(\sum_{j,j_1}^J \sigma_{i,jj_1}\right)^2 - \frac{2}{J}\sum_{j,j_1,j_2}^J \sigma_{i,jj_1}\sigma_{i,jj_2}\right],$$

$$\sigma_{AB} = \frac{1}{I(J-1)}\sum_{i=1}^I\sum_j^J \frac{\sigma_{i,jj}}{n_i} - \frac{1}{IJ(J-1)}\sum_{i=1}^I\sum_{j,j_1}^J \frac{\sigma_{i,jj_1}}{n_i},$$

provided that τ_{AB} *and* σ_{AB} *are bounded away from* 0 *and* ∞.

2.3.3 Tests for no time or treatment by time interaction effects from Bathke et. al (2008)

a. Test for no time effect

To test the null hypothesis of time effect in model 2.2.1, H0: $\beta_j = 0$, j =1, …, b, a test statistic given in is Bathke et. al (2008) is

$$F_B = \frac{N\bar{x}'_{..} P_b \bar{x}_{..}}{tr(P_b \hat{\Sigma}))}$$

where

$\bar{x}_{..} = X'(a^{-1}1_n \otimes n^{-1}1_n)$ is the vector of mean of the response variable, and

$$\hat{\Sigma} = \frac{1}{a(n-1)}[X'(I_a \otimes P_n)X]$$

.

Note that, we denoted the type I error rates from this test statistic as "Arne" for later simulation study of time effect.

b. Tests for no interaction effect

To test the null hypothesis of no interaction effects in model 2.2.1, H0: $\gamma_{ij} = 0$, where i=1, …, a and j =1, …, b, a test statistic given in Bathke et. al (2008) is

$$F_{AB} = \frac{N\bar{x}_{\cdot}'(P_a \otimes P_b)\bar{x}_{\cdot}}{tr[(P_a \otimes P_b)\hat{V}]}$$

where,

$$\bar{x}_{\cdot} = (I_a \otimes \frac{1}{n} 1_a' I_b) X$$

$$\hat{V} = I_a \otimes X'(I_a \otimes \frac{1}{n-1} P_b)$$

We denote this test statistic as "Arne" for the simulation study of interaction effect in later chapters.

Since the asymptotic distribution of above test statistic of the standardized F ratio is independent of the underlying population distribution of the response variable when there are a large number of treatments, the above test statistic F_{AB} could be approximated by its normal theory counterpart with assuming a spherical covariance structure to F-distribution with numerator degrees of freedom df_1= (a-1) (t-1) and denominator degrees of freedom df_2= a (t-1) (n-1). In later chapters, we denote this F-distribution approximation as "Arne2" in our simulation study.

CHAPTER 3 - Simulation Results

In this section, we report our simulation studies carried out to compare the test statistics reviewed in Chapter 2 and some traditional tests. All calculations and simulations were conducted with R (2.9.2) programming. All calculations for GEE are based on R package "geepack" those for LME are based on commands in package "nlme". All simulations reported here are based on 2000 runs per setting.

At first, we report some simulation study to evaluate type I error rates for random data generated from several distributions, such as normal distribution, gamma distribution, Poisson distribution, and beta distribution. Secondly, we use the power analysis to compare above methods with data generated from mixture distributions.

The datasets under the null hypothesis for normal distribution are based on standard normal; those for Gamma distribution have shape and scale parameters 1; those for Poisson distribution have mean of 4, those for beta distribution have shape and scale parameters 3 and 3. All datasets that we used for simulation study were generated based on the model as

$$X_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}$$

with an AR(1) correlation structure with autocorrelation ρ=0.5. Note that this simple correlation structure is to the advantage of the classical methods in that the large number of treatments basically provides additional observations to estimate the single common parameter ρ. That is, even though we are considering the case of a large number of treatments with small replications, the data for different treatments generated under the null hypothesis actually can be pooled together to estimate the common parameter ρ leading to a large sample size setting.

The next three subsections will present the simulation results. For a type I error estimate at level 0.05, if the estimate is outside of the 95% 'confidence' interval centered at 0.05, i.e., $0.05 \pm (0.05 \times 0.95/2000)^{0.5} \times 1.96 = (0.0404, 0.0596)$, the estimate is either conservative or liberal. We mark the estimated value with red color for liberal estimate and blue color for conservative estimate in all the tables presenting type I error estimate.

3.1 Simulation studies for treatment effaces

In this subsection, we compare the nonparametric test statistics by Zhang (2008), Harrar & Bathke(2008), and Bathke & Harrar(2008), and some traditional methods, such as linear mixed-effects models (LME), generalized estimating equations (GEE), Wilks' lambda, Hotelling-Lawley, and Pillai's multivariate tests by simulation study.

For all α-level simulation studies under the null hypotheses, we considered data generated with number of treatment a=10, 20, 40, 50, 100, 200, 500, the number of time point b=3, 8, and the number of replication n= 5 for balanced case or first 3 treatments with 3 replications and rest with 5 replications for unbalanced case.

First, we examined the tests for null hypotheses of no treatment effects For balanced case. For each distribution setting, all dataset under the null were generated using the same mean for all treatments (a=10, 20, 40, 50, 100, 200, 500) at all time points (t = 3, 8).

Tables 3-1 to 3-4 gave the type I error estimates with correlation ρ=0.5. At level α =0.05, the error rates of the test by Zhang (2008), are close to 0.05 for all datasets generated by normal, gamma, Poisson, and beta distributions when the number of treatments is 40 or higher and the number of time points being 3 or 8. When the number of treatments is less than 40, Zhang (2008) test is liberal. This is because Zhang's test is designed for large number of treatments.

For data from normal distribution, Hotelling-Lawley test also has liberal type I error estimate. This may indicate that Hotelling-Lawley requires large sample size for good performance. When there are 8 time points and 500 treatments, all tests except Zhang (2008) become conservative.

For data from gamma distribution, the multivariate tests and their nonparametric analogs are conservative when there are 40 treatments if the number of time points is 3. Similar phenomenon happens when the number of time points is 8 but when there are 200 time points. For data from Poisson distribution, we also observed similar pattern except that the estimates are less conservative when the number of time points is 3.

For data from beta distribution, all multivariate tests perform well but GEE and LME may be liberal in some cases.

No. time \ No. treatment		trt=10	trt=20	trt=40	trt=50	trt=100	trt=200	trt=500
t=3	Pillai	0.053	0.045	0.039	0.059	0.042	0.047	0.053
	Hotelling.Lawley	0.064	0.042	0.044	0.060	0.044	0.046	0.055
	Wilks	0.060	0.049	0.042	0.061	0.044	0.047	0.055
	Bathke.Pillai	0.051	0.042	0.039	0.058	0.040	0.046	0.052
	Bathke.Hotelling	0.056	0.048	0.043	0.058	0.043	0.046	0.055
	Zhang (2008)	0.084	0.062	0.051	0.051	0.052	0.051	0.050
	GEE	0.063	0.056	0.059	0.050	0.046	0.047	0.039
	LME	0.056	0.052	0.055	0.050	0.045	0.049	0.046
t=8	Pillai	0.038	0.043	0.051	0.042	0.051	0.056	0.031
	Hotelling.Lawley	0.075	0.062	0.059	0.048	0.054	0.056	0.031
	Wilks	0.053	0.054	0.052	0.044	0.053	0.055	0.031
	Bathke.Pillai	0.042	0.043	0.048	0.041	0.047	0.055	0.031
	Bathke.Hotelling	0.042	0.054	0.051	0.046	0.051	0.054	0.031
	Zhang (2008)	0.080	0.070	0.063	0.053	0.050	0.054	0.049
	GEE	0.063	0.055	0.050	0.053	0.049	0.067	0.031
	LME	0.084	0.052	0.056	0.057	0.050	0.071	0.036

Table 3-1 Estimated type I error rate of the test for null hypotheses with no treatment effect at α =0.05. The data generated by standard normal distribution with correlation ρ =0.5. The number of replications was 5.

No. time \ No. treatment		trt=10	trt=20	trt=40	trt=50	trt=100	trt=200	trt=500
t=3	Pillai	0.044	0.044	0.035	0.042	0.058	0.050	0.045
	Hotelling.Lawley	0.054	0.055	0.038	0.043	0.060	0.050	0.045
	Wilks	0.049	0.049	0.036	0.044	0.059	0.050	0.044
	Bathke.Pillai	0.044	0.042	0.034	0.049	0.056	0.046	0.043
	Bathke.Hotelling	0.049	0.049	0.037	0.050	0.058	0.050	0.044
	Zhang (2008)	0.057	0.065	0.061	0.054	0.053	0.052	0.052
	GEE	0.062	0.054	0.049	0.054	0.051	0.057	0.051
	LME	0.051	0.050	0.048	0.051	0.047	0.062	0.053
t=8	Pillai	0.043	0.045	0.051	0.044	0.048	0.037	0.046
	Hotelling.Lawley	0.047	0.051	0.055	0.054	0.049	0.038	0.047
	Wilks	0.046	0.046	0.051	0.049	0.048	0.037	0.044
	Bathke.Pillai	0.043	0.044	0.051	0.044	0.046	0.036	0.046
	Bathke.Hotelling	0.046	0.048	0.054	0.049	0.048	0.037	0.046
	Zhang (2008)	0.062	0.060	0.054	0.057	0.053	0.051	0.052
	GEE	0.066	0.054	0.046	0.062	0.043	0.039	0.056
	LME	0.087	0.061	0.051	0.066	0.041	0.044	0.061

Table 3-2 Estimated type I error rate of the test for null hypotheses with no treatment effect at α =0.05. The data generated by gamma distribution with correlation ρ =0.5. The number of replications was 5.

No. time \ No. treatment		trt=10	trt=20	trt=40	trt=50	trt=100	trt=200	trt=500
t=3	Pillai	0.047	0.049	0.044	0.039	0.043	0.045	0.048
	Hotelling.Lawley	0.062	0.054	0.050	0.037	0.044	0.047	0.049
	Wilks	0.052	0.052	0.047	0.041	0.045	0.045	0.049
	Bathke.Pillai	0.044	0.047	0.042	0.039	0.041	0.043	0.047
	Bathke.Hotelling	0.051	0.050	0.047	0.038	0.043	0.047	0.049
	Zhang (2008)	0.078	0.072	0.055	0.047	0.050	0.056	0.042
	GEE	0.068	0.057	0.051	0.048	0.057	0.073	0.038
	LME	0.057	0.054	0.051	0.043	0.057	0.078	0.041
t=8	Pillai	0.044	0.041	0.040	0.047	0.042	0.037	0.041
	Hotelling.Lawley	0.045	0.043	0.042	0.054	0.044	0.037	0.042
	Wilks	0.047	0.043	0.042	0.048	0.043	0.038	0.045
	Bathke.Pillai	0.044	0.038	0.040	0.046	0.039	0.039	0.044
	Bathke.Hotelling	0.046	0.042	0.041	0.050	0.042	0.038	0.044
	Zhang (2008)	0.091	0.075	0.059	0.054	0.051	0.056	0.042
	GEE	0.052	0.051	0.059	0.047	0.058	0.045	0.031
	LME	0.077	0.064	0.063	0.046	0.058	0.048	0.041

Table 3-3 Estimated type I error rate of the test for null hypotheses with no treatment effect at α =0.05. The data generated by Poisson distribution with correlation ρ =0.5. The number of replications was 5.

No. time \ No. treatment		trt=10	trt=20	trt=40	trt=50	trt=100	trt=200	trt=500
t=3	Pillai	0.043	0.051	0.049	0.057	0.049	0.043	0.051
	Hotelling.Lawley	0.053	0.059	0.049	0.057	0.050	0.046	0.051
	Wilks	0.046	0.053	0.048	0.057	0.050	0.045	0.051
	Bathke.Pillai	0.040	0.049	0.048	0.053	0.047	0.042	0.051
	Bathke.Hotelling	0.042	0.053	0.048	0.055	0.049	0.046	0.051
	Zhang (2008)	0.086	0.065	0.054	0.059	0.054	0.055	0.045
	GEE	0.055	0.048	0.056	0.053	0.062	0.056	0.055
	LME	0.061	0.048	0.057	0.054	0.059	0.058	0.058
t=8	Pillai	0.054	0.046	0.051	0.051	0.049	0.049	0.051
	Hotelling.Lawley	0.063	0.057	0.052	0.055	0.052	0.051	0.051
	Wilks	0.059	0.050	0.051	0.053	0.052	0.050	0.051
	Bathke.Pillai	0.053	0.044	0.047	0.050	0.048	0.049	0.051
	Bathke.Hotelling	0.055	0.051	0.051	0.051	0.050	0.050	0.051
	Zhang (2008)	0.092	0.069	0.069	0.064	0.049	0.049	0.045
	GEE	0.065	0.062	0.048	0.064	0.051	0.046	0.072
	LME	0.072	0.068	0.047	0.063	0.044	0.042	0.063

Table 3-4 Estimated type I error rate of the test for null hypotheses with no treatment effect at α =0.05. The data generated by beta distribution with correlation ρ =0.5. The number of replications was 5.

Secondly, we examined the tests for null hypotheses of no treatment effects for unbalanced dataset. The number of replications at each time point for the first 3 treatments is 3, and that for all other treatments is 5. Same as the balance case, all datasets generated by same distribution used the same mean for all treatments (a=10, 20, 40, 50, 100, 200, 500) at all time points (t = 3, 8).

For the unbalance data, neither GEE nor LME can be conducted successfully when there are a large number of treatments. When we ran the simulations with the R packages "geepack "and "nlme", there were some warnings, such as

#Warning messages:

#1: In lme.formula(y ~ as.factor(trt) + as.factor(Time), data =

data.frame(data), :

Reached total allocation of 1023Mb: see help(memory.size)

#2: In lme.formula(y ~ as.factor(trt) + as.factor(Time), data =

data.frame(data), :

Reached total allocation of 1023Mb: see help(memory.size)

Therefore, in this subsection, we report simulation results without GEE and LME.

Tables 3-5 to 3-8 present the type I error estimates for the unequal case with auto correlation ρ=0.5. In this case, we observed that the type I error estimates for all tests tend to be liberal for data from normal or gamma distribution when the number of time points is 3 and the number of treatments is small with the exception of trt=10. The multivariate tests and their nonparametric analogs are conservative when the number of treatments is 100 and the number of time point is 3 for these two distributions.

For data from Poisson distribution, the multivariate tests are conservative when there are 8 time points and small number of treatments. On the other hand, Zhang (2008)'s test is too liberal to be applicable in such case.

For data from beta distribution, majority of type I error estimates are good except for a few cases that the estimates of some tests are liberal.

No. time \ No. treatment		trt=10	trt=20	trt=40	trt=50	trt=100	trt=200	trt=500
t=3	Pillai	0.038	0.064	0.062	0.055	0.021	0.037	0.040
	Hotelling.Lawley	0.037	0.060	0.064	0.054	0.029	0.041	0.043
	Wilks	0.036	0.062	0.065	0.057	0.026	0.040	0.040
	Bathke.Pillai	0.041	0.061	0.060	0.054	0.035	0.047	0.045
	Bathke.Hotelling	0.042	0.060	0.060	0.055	0.037	0.050	0.045
	Zhang (2008)	0.083	0.070	0.062	0.063	0.055	0.051	0.050
t=8	Pillai	0.041	0.037	0.052	0.049	0.044	0.058	0.062
	Hotelling.Lawley	0.053	0.041	0.053	0.049	0.057	0.063	0.063
	Wilks	0.049	0.039	0.054	0.051	0.050	0.061	0.063
	Bathke.Pillai	0.051	0.045	0.052	0.050	0.053	0.055	0.057
	Bathke.Hotelling	0.051	0.046	0.053	0.052	0.053	0.057	0.060
	Zhang (2008)	0.080	0.068	0.063	0.057	0.051	0.052	0.051

Table 3-5 Estimated type I error rate of the test for null hypotheses with no treatment effect at α =0.05. The data generated by standard normal distribution with correlation ρ =0.5. Each of the first 3 treatments has 3 replications, and the others have 5 replications.

No. time \ No. treatment		trt=10	trt=20	trt=40	trt=50	trt=100	trt=200	trt=500
t=3	Pillai	0.041	0.072	0.065	0.063	0.024	0.045	0.043
	Hotelling.Lawley	0.040	0.068	0.067	0.062	0.032	0.049	0.046
	Wilks	0.039	0.070	0.068	0.065	0.029	0.048	0.043
	Bathke.Pillai	0.044	0.069	0.063	0.062	0.038	0.055	0.048
	Bathke.Hotelling	0.045	0.068	0.063	0.063	0.040	0.058	0.048
	Zhang (2008)	0.059	0.057	0.061	0.054	0.051	0.053	0.050
t=8	Pillai	0.049	0.040	0.060	0.052	0.052	0.059	0.063
	Hotelling.Lawley	0.061	0.044	0.061	0.052	0.065	0.064	0.064
	Wilks	0.057	0.042	0.062	0.054	0.058	0.062	0.064
	Bathke.Pillai	0.059	0.048	0.060	0.053	0.061	0.056	0.058
	Bathke.Hotelling	0.059	0.049	0.061	0.055	0.061	0.058	0.061
	Zhang (2008)	0.125	0.011	0.085	0.071	0.053	0.049	0.050

Table 3-6 Estimated type I error rate of the test for null hypotheses with no treatment effect at α =0.05. The data generated by gamma distribution with correlation ρ =0.5. Each of the first 3 treatments has 3 replications, and the others have 5 replications.

No. time \ No. treatment		trt=10	trt=20	trt=40	trt=50	trt=100	trt=200	trt=500
t=3	Pillai	0.037	0.045	0.048	0.038	0.042	0.048	0.047
	Hotelling.Lawley	0.048	0.049	0.054	0.036	0.043	0.051	0.049
	Wilks	0.040	0.047	0.050	0.040	0.044	0.049	0.048
	Bathke.Pillai	0.035	0.043	0.045	0.038	0.040	0.047	0.047
	Bathke.Hotelling	0.040	0.045	0.050	0.037	0.042	0.050	0.049
	Zhang (2008)	0.061	0.059	0.054	0.052	0.580	0.051	0.048
t=8	Pillai	0.035	0.037	0.043	0.045	0.041	0.039	0.042
	Hotelling.Lawley	0.035	0.039	0.045	0.052	0.043	0.039	0.043
	Wilks	0.037	0.039	0.045	0.046	0.042	0.040	0.045
	Bathke.Pillai	0.035	0.034	0.043	0.044	0.038	0.042	0.045
	Bathke.Hotelling	0.036	0.038	0.044	0.048	0.041	0.040	0.044
	Zhang (2008)	0.215	0.116	0.074	0.068	0.053	0.049	0.046

Table 3-7 Estimated type I error rate of the test for null hypotheses with no treatment effect at α =0.05. The data generated by Poisson distribution with correlation ρ =0.5. Each of the first 3 treatments has 3 replications, and the others have 5 replications.

No. time \ No. treatment		trt=10	trt=20	trt=40	trt=50	trt=100	trt=200	trt=500
t=3	Pillai	0.047	0.054	0.050	0.061	0.052	0.042	0.054
	Hotelling.Lawley	0.058	0.063	0.050	0.061	0.053	0.044	0.053
	Wilks	0.050	0.056	0.049	0.061	0.053	0.044	0.054
	Bathke.Pillai	0.044	0.052	0.049	0.056	0.050	0.041	0.053
	Bathke.Hotelling	0.047	0.056	0.048	0.058	0.052	0.044	0.053
	Zhang (2008)	0.086	0.065	0.054	0.059	0.054	0.055	0.045
t=8	Pillai	0.060	0.049	0.052	0.052	0.052	0.047	0.044
	Hotelling.Lawley	0.069	0.061	0.053	0.056	0.055	0.049	0.043
	Wilks	0.065	0.053	0.052	0.055	0.056	0.048	0.044
	Bathke.Pillai	0.059	0.047	0.048	0.051	0.051	0.047	0.043
	Bathke.Hotelling	0.060	0.054	0.052	0.052	0.053	0.048	0.043
	Zhang (2008)	0.092	0.069	0.069	0.064	0.049	0.049	0.045

Table 3-8 Estimated type I error rate of the test for null hypotheses with no treatment effect at α =0.05. The data generated by beta distribution with correlation ρ =0.5. Each of the first 3 treatments has 3 replications, and the others have 5 replications.

3.2 Simulation study for time effects

In this subsection, we report simulations for null hypothesis of no time effects. Similar with the simulation study for the treatment effects, we randomly generated data from normal, gamma, Poisson, and beta distribution for balanced design and unbalanced design. However, the multivariate tests and their nonparametric analogs can not test for the time effect. Therefore, we only compare tests of Zhang (2008) with GEE, LME, and Bathke et al. (2008) [5] (denoted as Arne in all tables thereafter) in this subsection. The test in Bathke et al. (2008) that we will compare is a small sample size adjustment of ANOVA type statistics.

First, we consider type I error estimates in the balanced case. All datasets were randomly generated under the null hypotheses of no time effect. Five replications were generated for each treatment exactly the same as in the balanced case for the test of no treatment effect.

The type I error rates at α =0.05 are shown in Tables 3-9 to 3-12. It is clear that the chi-square test of Zhang (2008) is very liberal when the number of time point is 8 and the number of treatment is not large ($\leq$ 50 for normal and Poisson distribution; $\leq$ 20 for beta distribut ion, $\leq$ 10 for gamma distribution). Bathke et al. (2008)'s small sample size adjustment works very well for normal and gamma distributed data. For other tests, it is a general pattern that the estimate is liberal when the number of treatments is 10 or 20. We remark that is still an indication of large sample size requirement since all data under the null have identical distributions for all treatments.

No. time \ No. treatment		trt=10	trt=20	trt=40	trt=50	trt=100	trt=200	trt=500
t=3	Zhang (2008)	0.061	0.059	0.058	0.051	0.058	0.052	0.052
	Arne	0.048	0.055	0.056	0.051	0.055	0.055	0.051
	GEE	0.057	0.061	0.054	0.050	0.049	0.049	0.038
	LME	0.046	0.060	0.049	0.050	0.046	0.045	0.037
t=8	Zhang (2008)	0.140	0.086	0.066	0.061	0.056	0.053	0.051
	Arne	0.046	0.047	0.048	0.050	0.052	0.049	0.040
	GEE	0.071	0.058	0.047	0.046	0.043	0.050	0.036
	LME	0.096	0.065	0.052	0.053	0.045	0.051	0.044

Table 3-9 Estimated type I error rate of the test for null hypotheses with no time effect at α =0.05. The data generated by standard normal distribution with correlation ρ =0.5. The number of replications is 5.

No. time \ No. treatment		trt=10	trt=20	trt=40	trt=50	trt=100	trt=200	trt=500
t=3	Zhang (2008)	0.071	0.060	0.053	0.052	0.053	0.044	0.053
	Arne	0.051	0.050	0.047	0.044	0.050	0.041	0.053
	GEE	0.056	0.058	0.052	0.057	0.060	0.047	0.044
	LME	0.051	0.052	0.052	0.056	0.053	0.049	0.040
t=8	Zhang (2008)	0.068	0.055	0.051	0.052	0.052	0.054	0.056
	Arne	0.055	0.047	0.044	0.047	0.048	0.042	0.050
	GEE	0.074	0.054	0.058	0.056	0.054	0.046	0.042
	LME	0.098	0.055	0.060	0.065	0.055	0.042	0.037

Table 3-10 Estimated type I error rate of the test for null hypotheses with no time effect at α =0.05. The data generated by gamma distribution with correlation ρ =0.5. The number of replications is 5.

No. time \ No. treatment		trt=10	trt=20	trt=40	trt=50	trt=100	trt=200	trt=500
t=3	Zhang (2008)	0.078	0.064	0.060	0.051	0.055	0.044	0.050
	Arne	0.064	0.056	0.056	0.052	0.050	0.043	0.051
	GEE	0.068	0.061	0.056	0.052	0.057	0.049	0.037
	LME	0.057	0.054	0.056	0.045	0.042	0.053	0.035
t=8	Zhang (2008)	0.087	0.070	0.067	0.061	0.048	0.052	0.051
	Arne	0.063	0.055	0.039	0.054	0.036	0.053	0.046
	GEE	0.058	0.060	0.054	0.050	0.066	0.058	0.031
	LME	0.080	0.067	0.061	0.052	0.062	0.062	0.051

Table 3-11 Estimated type I error rate of the test for null hypotheses with no time effect at α =0.05. The data generated by Poisson distribution with correlation ρ =0.5. The number of replications is 5.

No. time \ No. treatment		trt=10	trt=20	trt=40	trt=50	trt=100	trt=200	trt=500
t=3	Zhang (2008)	0.074	0.060	0.057	0.050	0.058	0.050	0.048
	Arne	0.060	0.053	0.047	0.047	0.058	0.046	0.048
	GEE	0.068	0.051	0.054	0.056	0.070	0.061	0.045
	LME	0.059	0.049	0.051	0.052	0.061	0.059	0.045
t=8	Zhang (2008)	0.079	0.061	0.059	0.055	0.049	0.046	0.048
	Arne	0.051	0.061	0.044	0.042	0.046	0.045	0.043
	GEE	0.073	0.062	0.043	0.055	0.051	0.077	0.063
	LME	0.097	0.067	0.051	0.060	0.041	0.081	0.054

Table 3-12 Estimated type I error rate of the test for null hypotheses with no time effect at α =0.05. The data generated by beta distribution with correlation ρ =0.5. The number of replications is 5.

Next, we report the test results for null hypotheses of no time effects for unbalanced design. The data generation is same as those for the unbalanced case in the test of no main treatment effect. That is, 3 replications were generated for the first 3 treatments and 5 replications for all other treatments. Similar to the test of treatment effects with unbalanced data, GEE and LME fail to work when there were a large number of treatments.

Tables 3-13 to 3-16 gave the type I error rates for the test of no main time effect for unbalanced design. The conclusions are similar to the balanced case.

No. time \ No. treatment		trt=10	trt=20	trt=40	trt=50	trt=100	trt=200	trt=500
t=3	Zhang (2008)	0.062	0.060	0.053	0.055	0.051	0.044	0.049
	Arne	0.044	0.062	0.056	0.057	0.039	0.041	0.047
t=8	Zhang (2008)	0.093	0.070	0.066	0.059	0.052	0.049	0.053
	Arne	0.054	0.044	0.055	0.049	0.049	0.053	0.059

Table 3-13 Estimated type I error rate of the test for null hypotheses with no time effect at α =0.05. The data generated by standard normal distribution with correlation ρ =0.5. Each of the first 3 treatments has 3 replications, and the others have 5 replications.

No. time \ No. treatment		trt=10	trt=20	trt=40	trt=50	trt=100	trt=200	trt=500
t=3	Zhang (2008)	0.071	0.062	0.059	0.060	0.054	0.049	0.049
	Arne	0.046	0.072	0.066	0.066	0.040	0.060	0.049
t=8	Zhang (2008)	0.110	0.069	0.072	0.060	0.054	0.052	0.043
	Arne	0.061	0.050	0.064	0.057	0.064	0.060	0.064

Table 3-14 Estimated type I error rate of the test for null hypotheses with no time effect at α =0.05. The data generated by gamma distribution with correlation ρ =0.5. Each of the first 3 treatments has 3 replications, and the others have 5 replications.

No. time \ No. treatment		trt=10	trt=20	trt=40	trt=50	trt=100	trt=200	trt=500
t=3	Zhang (2008)	0.073	0.065	0.056	0.061	0.052	0.049	0.049
	Arne	0.039	0.048	0.051	0.043	0.046	0.053	0.053
t=8	Zhang (2008)	0.093	0.079	0.067	0.060	0.053	0.047	0.043
	Arne	0.036	0.035	0.045	0.045	0.039	0.043	0.046

Table 3-15 Estimated type I error rate of the test for null hypotheses with no time effect at α =0.05. The data generated by Poisson distribution with correlation ρ =0.5. Each of the first 3 treatments has 3 replications, and the others have 5 replications.

No. time \ No. treatment		trt=10	trt=20	trt=40	trt=50	trt=100	trt=200	trt=500
t=3	Zhang (2008)	0.074	0.057	0.048	0.052	0.048	0.049	0.041
	Arne	0.060	0.053	0.047	0.047	0.058	0.046	0.048
t=8	Zhang (2008)	0.078	0.060	0.060	0.056	0.044	0.044	0.041
	Arne	0.050	0.060	0.045	0.042	0.045	0.048	0.048

Table 3-16 Estimated type I error rate of the test for null hypotheses with no time effect at α =0.05. The data generated by beta distribution with correlation ρ =0.5. Each of the first 3 treatments has 3 replications, and the others have 5 replications.

3.3 Simulation study for treatment and time point interaction effects

In this subsection, we report simulations for null hypotheses of no interaction effects of treatment and time. Similar to the simulation study for main treatment effects and time effects, we randomly generated data from normal, gamma, Poisson, and beta distribution with equal number of replications or unequal replications. And for the same reason as explained in section 3.2, in this part, we only compare tests of Zhang (2008) with GEE, LME, Bathke et al. (2008) [5] (that was denoted as Arne, Arne2).

Tables 3-17 to 3-20 gave type I error estimates for the test of no interaction effect at α =0.05. The performance of Zhang (2008) has similar patter to the test of no main treatment effect, i.e., the type I error estimate is liberal for small number of treatments in all distribution settings. GEE and LME (especially LME) become liberal in some settings under non-normal distributions. It is very clear that Arne by Bathke et al. (2008) [5] becomes more and more conservative as the number of treatments increases.

No. time \ No. treatment		trt=10	trt=20	trt=40	trt=50	trt=100	trt=200	trt=500
t=3	Zhang (2008)	0.080	0.063	0.054	0.060	0.055	0.048	0.043
	Arne	0.048	0.033	0.028	0.029	0.022	0.022	0.016
	Arne2	0.051	0.048	0.044	0.047	0.046	0.051	0.048
	GEE	0.059	0.056	0.050	0.047	0.045	0.052	0.078
	LME	0.055	0.055	0.053	0.048	0.042	0.047	0.098
t=8	Zhang (2008)	0.073	0.059	0.055	0.048	0.056	0.057	0.051
	Arne	0.011	0.010	0.004	0.002	0.004	0.003	0.000
	Arne2	0.051	0.050	0.048	0.042	0.054	0.051	0.040
	GEE	0.063	0.047	0.048	0.050	0.429	0.057	0.062
	LME	0.093	0.057	0.054	0.050	0.045	0.064	0.071

Table 3-17 Estimated type I error rate of the test for null hypotheses with no interaction effect of treatment and time at α =0.05. The data generated by standard normal distribution with correlation ρ =0.5. The number of replications is 5.

No. time \ No. treatment		trt=10	trt=20	trt=40	trt=50	trt=100	trt=200	trt=500
t=3	Zhang (2008)	0.057	0.066	0.064	0.056	0.052	0.053	0.045
	Arne	0.004	0.023	0.021	0.016	0.019	0.009	0.007
	Arne2	0.048	0.046	0.050	0.049	0.054	0.047	0.045
	GEE	0.055	0.056	0.051	0.053	0.051	0.055	0.042
	LME	0.046	0.053	0.053	0.053	0.046	0.057	0.038
t=8	Zhang (2008)	0.066	0.062	0.052	0.051	0.048	0.050	0.043
	Arne	0.003	0.002	0.019	0.004	0.007	0.001	0.003
	Arne2	0.051	0.044	0.047	0.048	0.048	0.039	0.043
	GEE	0.063	0.047	0.055	0.060	0.056	0.050	0.065
	LME	0.088	0.057	0.061	0.064	0.055	0.046	0.056

Table 3-18 Estimated type I error rate of the test for null hypotheses with no interaction effect of treatment and time at α =0.05. The data generated by gamma distribution with correlation ρ =0.5. The number of replications is 5.

No. time \ No. treatment		trt=10	trt=20	trt=40	trt=50	trt=100	trt=200	trt=500
t=3	Zhang (2008)	0.076	0.053	0.058	0.050	0.043	0.049	0.060
	Arne	0.048	0.030	0.023	0.011	0.019	0.017	0.022
	Arne2	0.054	0.043	0.044	0.045	0.042	0.051	0.053
	GEE	0.069	0.058	0.054	0.049	0.059	0.063	0.072
	LME	0.057	0.054	0.051	0.052	0.051	0.063	0.083
t=8	Zhang (2008)	0.068	0.054	0.051	0.050	0.050	0.054	0.048
	Arne	0.011	0.032	0.007	0.004	0.001	0.020	0.033
	Arne2	0.055	0.049	0.046	0.048	0.051	0.047	0.049
	GEE	0.056	0.052	0.059	0.044	0.057	0.060	0.061
	LME	0.078	0.060	0.067	0.052	0.061	0.059	0.041

Table 3-19 Estimated type I error rate of the test for null hypotheses with no interaction effect of treatment and time at α =0.05. The data generated by Poisson distribution with correlation ρ =0.5. The number of replications is 5.

No. time \ No. treatment		trt=10	trt=20	trt=40	trt=50	trt=100	trt=200	trt=500
t=3	Zhang (2008)	0.079	0.059	0.062	0.053	0.048	0.043	0.051
	Arne	0.049	0.034	0.030	0.026	0.024	0.023	0.017
	Arne2	0.054	0.047	0.057	0.053	0.058	0.051	0.054
	GEE	0.055	0.053	0.052	0.056	0.060	0.055	0.044
	LME	0.061	0.053	0.053	0.058	0.055	0.056	0.045
t=8	Zhang (2008)	0.090	0.060	0.056	0.057	0.041	0.052	0.051
	Arne	0.056	0.038	0.031	0.028	0.022	0.017	0.017
	Arne2	0.063	0.051	0.055	0.051	0.051	0.053	0.054
	GEE	0.061	0.063	0.048	0.057	0.050	0.069	0.081
	LME	0.092	0.067	0.055	0.066	0.050	0.066	0.072

Table 3-20 Estimated type I error rate of the test for null hypotheses with no interaction effect of treatment and time at α =0.05. The data generated by beta distribution with correlation ρ =0.5. The number of replications is 5.

Secondly, simulation study was conducted to test the interaction effect under the null hypotheses for unbalanced design. Data were also generated under the same settings as those for the test of no main treatment effect. Again, GEE and LME are removed for comparison for unbalanced data. Tables 3-21 to 3-24 showed that the conclusions were mostly same with the balanced case.

No. time \ No. treatment		trt=10	trt=20	trt=40	trt=50	trt=100	trt=200	trt=500
t=3	Zhang (2008)	0.071	0.072	0.061	0.050	0.051	0.060	0.052
	Arne	0.011	0.020	0.022	0.031	0.026	0.019	0.017
	Arne2	0.053	0.046	0.049	0.058	0.052	0.050	0.051
t=8	Zhang (2008)	0.073	0.059	0.055	0.048	0.056	0.057	0.051
	Arne	0.009	0.013	0.005	0.002	0.000	0.004	0.001
	Arne2	0.052	0.053	0.050	0.047	0.053	0.052	0.044

Table 3-21 Estimated type I error rate of the test for null hypotheses with no interaction effect at α =0.05. The data generated by standard normal distribution with correlation ρ =0.5. Each of the first 3 treatments has 3 replications, and the others have 5 replications.

No. time \ No. treatment		trt=10	trt=20	trt=40	trt=50	trt=100	trt=200	trt=500
t=3	Zhang (2008)	0.057	0.066	0.064	0.056	0.052	0.053	0.045
	Arne	0.022	0.015	0.012	0.030	0.021	0.018	0.014
	Arne2	0.040	0.061	0.056	0.056	0.036	0.052	0.043
t=8	Zhang (2008)	0.066	0.062	0.052	0.051	0.048	0.050	0.043
	Arne	0.003	0.006	0.003	0.004	0.003	0.003	0.001
	Arne2	0.053	0.044	0.054	0.049	0.054	0.051	0.054

Table 3-22 Estimated type I error rate of the test for null hypotheses with no interaction effect at α =0.05. The data generated by gamma distribution with correlation ρ =0.5. Each of the first 3 treatments has 3 replications, and the others have 5 replications.

No. time \ No. treatment		trt=10	trt=20	trt=40	trt=50	trt=100	trt=200	trt=500
t=3	Zhang (2008)	0.067	0.055	0.061	0.052	0.054	0.043	0.051
	Arne	0.016	0.014	0.009	0.027	0.015	0.016	0.010
	Arne2	0.060	0.058	0.062	0.055	0.052	0.480	0.051
t=8	Zhang (2008)	0.075	0.062	0.058	0.052	0.047	0.051	0.049
	Arne	0.003	0.005	0.003	0.004	0.003	0.003	0.001
	Arne2	0.068	0.051	0.047	0.052	0.051	0.044	0.051

Table 3-23 Estimated type I error rate of the test for null hypotheses with no interaction effect at α =0.05. The data generated by Poisson distribution with correlation ρ =0.5. Each of the first 3 treatments has 3 replications, and the others have 5 replications.

No. time \ No. treatment		trt=10	trt=20	trt=40	trt=50	trt=100	trt=200	trt=500
t=3	Zhang (2008)	0.066	0.051	0.043	0.046	0.043	0.044	0.046
	Arne	0.013	0.011	0.012	0.025	0.013	0.014	0.007
	Arne2	0.061	0.059	0.063	0.056	0.054	0.055	0.048
t=8	Zhang (2008)	0.081	0.066	0.066	0.062	0.052	0.052	0.049
	Arne	0.006	0.008	0.006	0.007	0.006	0.000	0.004
	Arne2	0.069	0.053	0.048	0.054	0.053	0.039	0.053

Table 3-24 Estimated type I error rate of the test for null hypotheses with no interaction effect at α =0.05. The data generated by beta distribution with correlation ρ =0.5. Each of the first 3 treatments has 3 replications, and the others have 5 replications.

3.4 Power analysis study

In this subsection, we compare the estimated power for the tests of Zhang (2008) with tests by Harrar & Bathke (2008), Bathke & Harrar(2008) and some traditional methods, such as LME, GEE, Wilks' lambda, Hotelling-Lawley, and Pillai's multivariate tests. We generated the data with auto correlation ρ =0.5. The power analysis reported here is based on 2000 runs per setting.

Figure 3.4.a gives the estimated power curves of the tests of Zhang (2008), Harrar & Bathke(2008), Bathke & Harrar(2008), LME, GEE, Wilks' lambda, Hotelling-Lawley, and Pillai's multivariate tests in balanced design with 500 treatments, 3 time pints, and 5 replications. The data for the i^{th} treatment were generated from one of the two distributions: normal distribution with mean $\mu_i=3+(b/i)\tau$ if the $i\leq 100$; gamma distribution with shape parameter $\mu_i=\mu+(b/i)\tau$ and scale parameter 1 if $i>100$, where b =3 is the number of time points, for τ=1, 10, 20, 30, 40, 50, 60, 70, 80, 90, and 100.

The tests of Zhang (2008) has the highest power when τ is greater than 20. At τ=60, the estimated power of the nonparametric analog of Phillai' test of Harrar & Bathke(2008) is 66%, that of Hotelling-Lawley test of Harrar & Bathke(2008) is 77%, that of LME is 53%, that of GEE is 24% , that of Wilks' lambda is 35%, that of Hotelling-Lawley is 30%, and that of Phillai's multivariate test is 33%. The estimated powers of the tests of Harrar & Bathke(2008) and LME did not reach 1 until τ is 80. GEE, Hotelling-Lawley, and wilks' lambda had the lowest power.

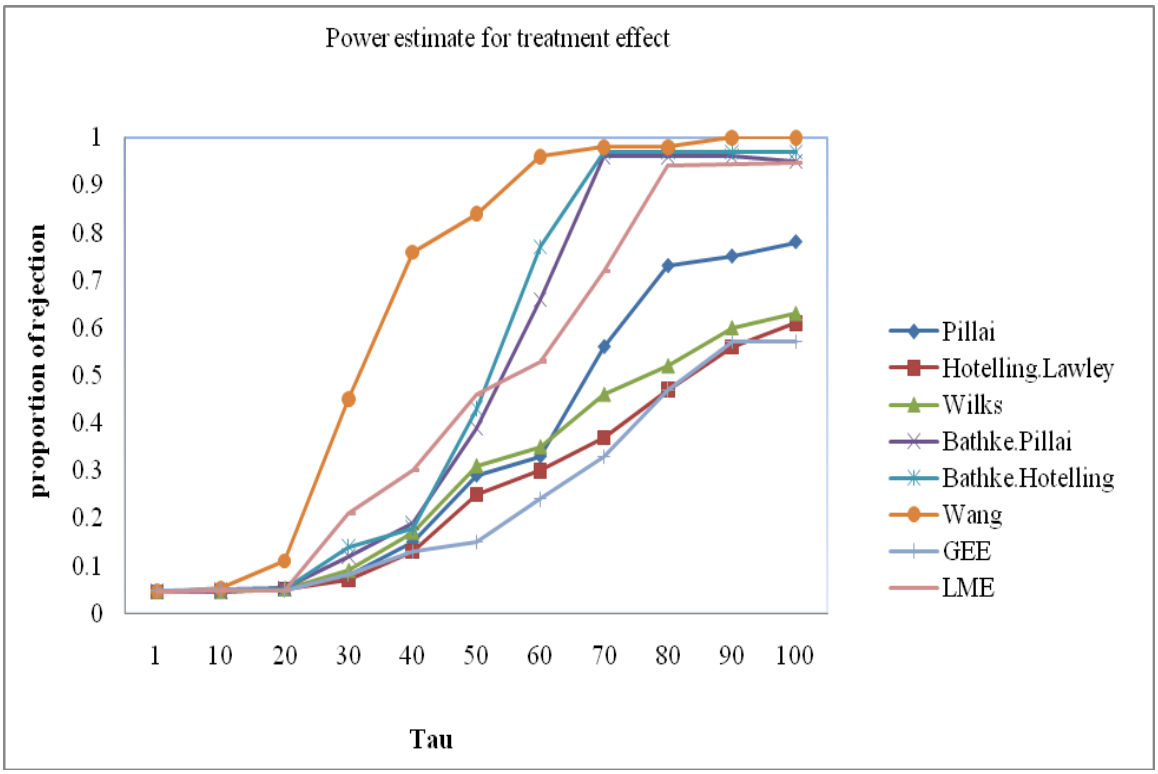

Figure 3-1 The estimated power curves at level 0.05 of treatment effects for balanced design with 500 treatments, 3 time points, and 5 replications.

Figure 3.4.b showed that power curves of test of Zhang (2008), test by Harrar & Bathke (2008), Wilks' lambda, Hotelling-Lawley, and Pillai's multivariate tests under unbalance design with 500 treatments, 3 time pints, and different replications for treatment, such as the first 3 had 3 replications, others had 5 replications. Since GEE and LME were not work for large number of treatment under unbalance data, there were no curves for them. The test of Zhang (2008) had the highest power when tau greater than 10. At tau was 100, the powers of Pillai and Hotelling-Lawley that both based on studies of Harrar & Bathke(2008), and Bathke & Harrar(2008) reached to 91% and 93.5%. At tau was 100, Power of Pillai's multivariate tests, Hotelling-Lawley, and wilks' lambda had the worst power that were lower than 70%.

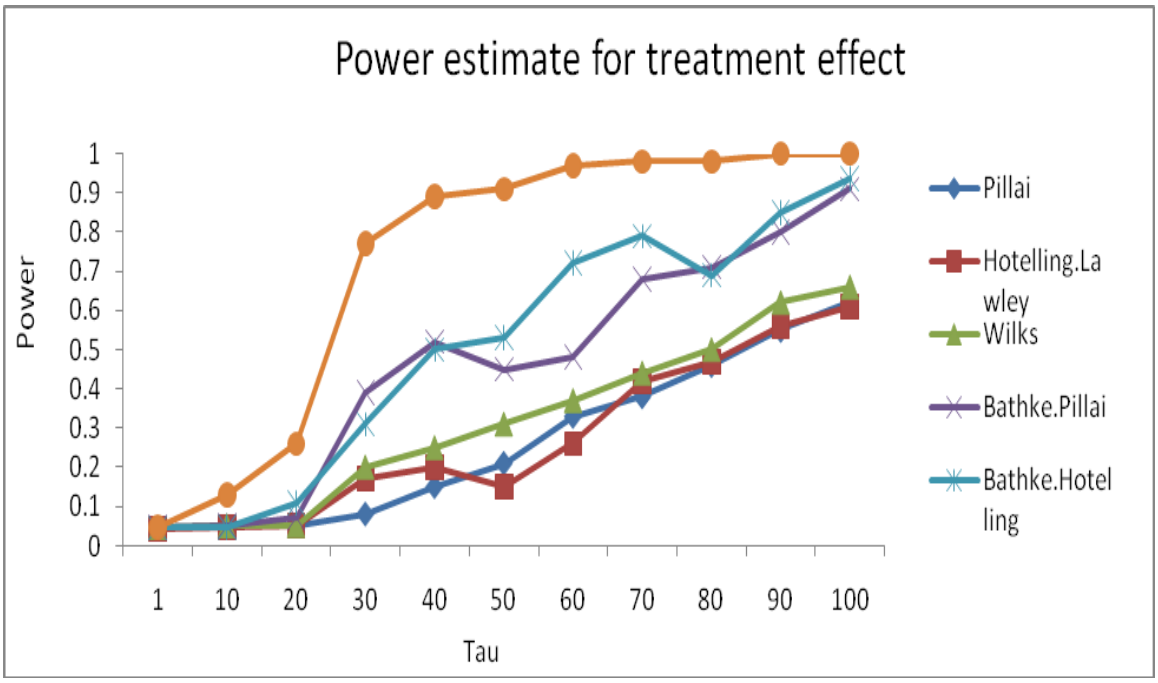

Figure 3-2 The power curves of treatment effect under unbalanced design with 500 treatments, 3 time points, and first 3 with 3 replications and others 5 replications.

Next, we report the power ananlysis for time effect for data generated under balanced design. The number of treatments is 500, number of time points is 3, and the number of replications for each treatment is 5. Similar to the power analysis of treatment effect, the data for treatment i was generated from one of the two distributions: normal distribution with mean $\mu_j=3+(j/a)\tau$ if $j \leq 2$; gamma distribution with shape parameter $\mu_j=3+\tau(j/a)$ and scale parameter 1 if $j>2$, where a is the number of treatments, j=1, 2, 3, τ=1, 10, 20, 30, 40, 50, 60, 70, 80, 90, and100. Only comparisons among tests of Zhang(2008), Bathke et al. (2008) [5], GEE, and LME are reported here.

Figure 3.4.c showed that the chi-square test of Zhang (2008) and the test of Bathke (2008) [5] have comparable estimated power that is higher than GEE and LME for large values of τ. GEE has the worst power among these four methods.

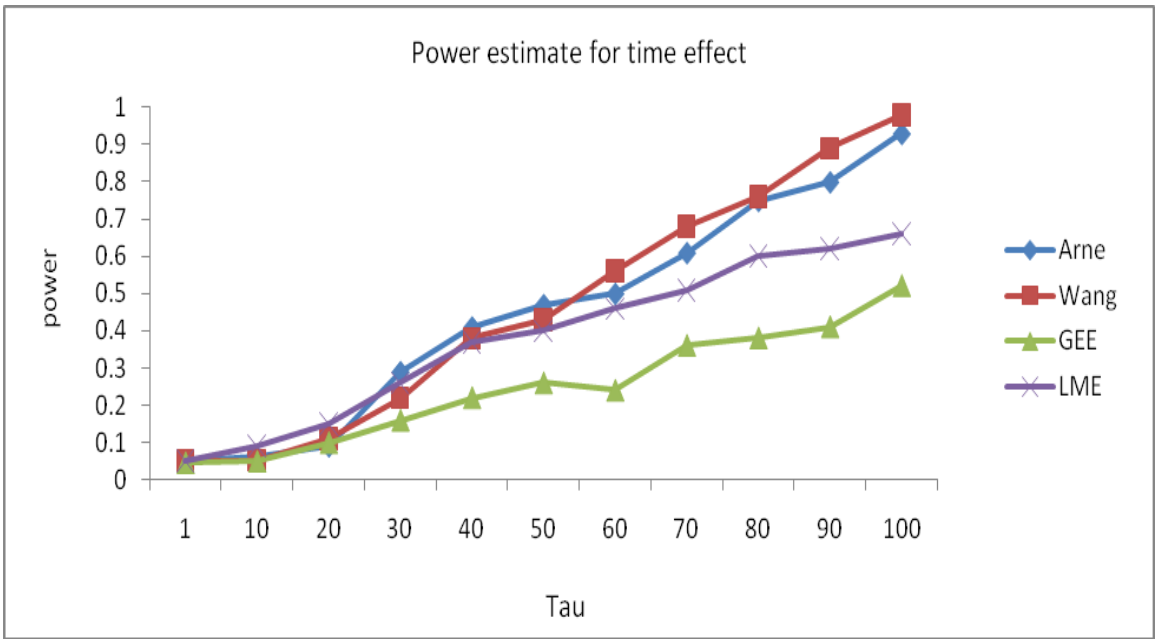

Figure 3-3 The power curves of time effect under balanced design with 500 treatments, 3 time points, and 5 replications.

CHAPTER 4 - Conclusion and recommendations

In this report, we reviewed and numerically compared several tests that are potentially applicable to high dimensional longitudinal data. For the test of no main treatments, potential tests include Zhang (2008), linear mixed-effects models (LME), generalized estimating equations (GEE), Wilks' lambda, Hotelling-Lawley, and Phillai's multivariate tests, and their nonparametric analogs by Harrar & Bathke (2008) and Bathke & Harrar(2008), under balanced and unbalanced datasets. For the test of no main time effect or treatment by time interaction effects, potential tests include Zhang (2008), linear mixed-effects models (LME), generalized estimating equations (GEE), and the tests in Bathke et al. (2009).

Our simulation studies suggest that the type I error estimates of the tests in Zhang (2008) converges to 0.05 as the number of treatments increases for both balanced and unbalanced designs for all distributions considered. The test statistics based on ratio of two quadratic forms are insensitive to the number of time points. However, the Wald-type test statistic in Zhang (2008) have empirical type I error converges to the true level very slowly as the number of treatments increases. In fact, when the number of time points is relatively medium compared to the number of treatments, the type I error of the Wald-type test can be very poor.

GEE and LME may be liberal for some cases and may be conservative for some other cases even though their type I error estimates are within the acceptable range majority of time. Their performance is less consistent across different distributions. Similar phenomenon is observed for the multivariate tests. The nonparametric analogs of the multivariate tests seem to have better type I errors and empirical power than the multivariate tests, but are not as powerful as the test of Zhang (2008) when the number of treatments is truly large (500). When the number of treatments is small or moderately large, the tests of LME, GEE, Wilks' lambda, Hotelling-Lawley, and Pillai's multivariate tests, Harrar & Bathke (2008) seem to be better than the test of Zhang (2008) for the auto-correlated data generated in this report in type I error estimate.

Based on the power analysis, GEE, LME, Wilks' lambda, Hotelling-Lawley, and Pillai's multivariate tests have lower power than tests of Zhang (2008) and Harrar & Bathke (2008) for detecting main treatment effects. For main time effects, the powers of Bathke et al. (2008) [5] and

Zhang (2008) are comparable and are both more powerful than the traditional tests in LME or GEE.

According to the results of the simulation studies, the traditional methods will perform well for the longitudinal data with small number of treatments. But when the number of treatments is large, the tests of Zhang (2008) should be more useful in identify significance effects in the data.

References

1. Arnold, S.F (1981). *The Theory of Linear Models and Multivariate Analysis*, New York, Wiley.
2. Bathke, A.C., Harrar, S.W., and Madden, H.L (2008). How to compare small multivariate samples using nonparametric tests. *Computaional Statistics and Data Analysis*, 52, 4951-4965.
3. Bathke, A.C. and Harrar, S.W. (2008). Nonparametric methods in multivariate factorial designs for large number of factor levels. *Journal of Statistical planning and Inference*, 138, 1,588-610.
4. Bathke, B.C., Harrar, S.W., Wang, H, Zhang, K, and Piepho, H. (2008). Series of randomized complete block experiments with non-normal data. Unpublished manuscript.
5. Brunner E., Domhof E., and Langer F. (2002). *Nonparametric analysis for longitudinal data in factorial experiment*, Wiley.
6. Gary A. Ballinger (2004). Using Generalized Estimating Equations for Longitudinal Data Analysis, *Organizational Research Methods*, 7, No. 2, 127-150.
7. Harrar, S.W. and Bathke, A.C. (2008a). A nonparametric version of the Bartlett-Nanda-Pillai Multivariate Test: Asymptotic, Approximations, and applications. *American Journal of Mathematical and Management Sciences*, in press.
8. Hedeker D. and Dibbons R. (2006). *Longitudinal Data Analysis*, Wiley.
9. Liang, K.Y. and Zeger, S.L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73, 1, 13-22.
10. Wang, H. and Akritas, M. (2009). Inference from heteroscedastic functional data. to appear in *Journal of Nonparametric Statistics*, DOI:10.1080/10485250903171621.
11. Wang, H., Higgins, J., and Blasi, D. (2009). Distribution-free tests for no effect of treatment in heteroscedastic functional data under both weak and long range dependence. to appear in *Statistics and Probability Letters*, DOI:10.1016/j.spl.2009.11.016.
12. Zhang, Ke (2008). Inference of nonparametric hypothesis testing on high dimensional longitudinal data and its application in DNA copy number variation and microarray data analysis. PhD. Dissertation, Kansas State University.

APPENDIX A - R codes for data analysis

In the Appendix, we provide the R functions that were used for the simulation study and for the real data analysis.

In the R functions for longitudinal data study, there were several input parameters to use, such as the input data that denote as "data", the amount of total treatments that denote as "a", the amount of total time points that denote as "b", the number of columns of input data that denote as "coln", and the vector of the number of replications for each treatment that denoted as "n".

For the study, the input data is a data matrix with 4 columns. The first column is the treatment, second column is the time point, third column is the replication, and the forth column is the response. Then, the response is denoted as X_{ijk} that is a response of the i^{th} treatment, the j^{th} time point, and the k^{th} replicate. Therefore, the data matrixes as the following format

$$
\begin{array}{cccc}
1 & 1 & 1 & x_{111} \\
1 & 1 & 2 & x_{112} \\
1 & 2 & 1 & x_{121} \\
1 & 2 & 2 & x_{122} \\
1 & 2 & 3 & x_{123} \\
2 & 1 & 1 & x_{211}
\end{array}
$$

R code for generate data:

A1: To generate data

```
#dataformat2 function takes argument data in format1 and converts to dataformat2
#### x_{ijk}, k=1, ..., n_i are the kth observation from the ith subject at time j.
# 1 1 1 x111
# 1 1 2 x112
# 1 2 1 x121
# 1 2 2 x122
# 1 2 3 x123

dataformat2= function(data){
        m=ncol(data)-2
        y=c(t(data[,-c(1,2)]))
```

```
        Time=rep(1:m, nrow(data))
        sub=kronecker( data[,1],rep(1, m))
        trt=kronecker( data[,2],rep(1, m))
        mydata= cbind(trt, Time, sub, y)
        }

f2=function(t,u) rnorm(t,u)
f3=function(t) rpois(t,4)
f4=function(t,u) rgamma(t,u)
f5=function(t) rbeta(t,3,3)

### R code for null hypotheses data
datagen=function(a,m,mn,mu,tau,sigmaj =runif(1, 1.2, 1.4) ){
rho=0.5
 covm=rho^abs(matrix(rep(seq(m), m), m, m)- matrix(rep(seq(m), m), m, m, byrow=T))
 tran=t(chol(covm) )   # transformation matrix. if X ~ N(0, I), then tran X ~ N(0, A) with
                # A being the cov matrix of AR(1), the covariance is sigma and the
                # correlation coeff is rho=exp(-1/m).
                # i.e. A= sigma(1  rho  rho^2 ... rho^(m-1)
                #              rho 1   rho   ... rho^(m-2)
                #                .....................
                #              rho^(m-1) rho^(m-2) ... rho)
### generate data and store it in format1
data=numeric()
now=0
for (j in 1:a){
        muj=mu
        fj=function(x, v) f2(x, v)
        for (i in 1:mn[j]){
                now=now+1
                data=rbind(data, c(i, j, muj*fj(m,muj)+ c(sigmaj*tran%*%rnorm(m))))
```

```
                colnames(data)=c("sub", "trt", paste("time", seq(m), sep=""))
                 }
                } #end of j
#above data is in format1
mydat=dataformat2(data)
}

#write.table(mydat, file=paste("dat", k, ".txt",sep=""), row.names = F, col.names = T)
list(dat.format1=data, mydat=mydat, ke.format=ke.format)
}
```

A2: Parts of R code for generated data for alterative hypotheses:

For treatment effect:

```
for (j in 1:a){
        muj=mu+(m/j)*tau
fj=function(x,v) f2(x,v)*(j<=100)+f4(x,v)*(j>100)
        for (i in 1:mn[j]){
        now=now+1
         data=rbind(data, c(i, j, muj*fj(m,muj)+ c(sigmaj*tran%*%rnorm(m))))
                        colnames(data)=c("sub", "trt", paste("time", seq(m), sep=""))
                         }
                } #end of j
```

For time effect:

```
for (j in 1:a){
for (k in 1:m){
        muj=mu+(k/a)*tau
fj=function(x,v) f2(x,v)*(k<=2)+f4(x,v)*(k>2)
        for (i in 1:mn[j]){
        now=now+1
         data=rbind(data, c(i, j, muj*fj(m,muj)+ c(sigmaj*tran%*%rnorm(m))))
```

```
                colnames(data)=c("sub", "trt", paste("time", seq(m), sep=""))
                }
        }#end of k
        }#end of j
```

A3: R code for calculation of the proposed method:

```
dot=function(xbar,n){
    a=length(n)
    result=numeric()
    for(i in 1:a) result=c(result,rep(xbar[i,],n[i]))
    result
    }

ncal=function(data){
    a=length(unique(data[,1]))
    b=length(unique(data[,2]))
    mn=numeric()
    for(i in 1:a) mn[i]=length(unique(data[data[,1]==i,3]))
    list(n=mn,a=a,b=b)
    }
### calculate the MSE

msecal=function(newdat,a,n,b){
    res2=0
    resv=0
    for(i in 1:a){
        temp=as.data.frame(newdat[newdat[,1]==i,])
        tempfunc=function(x){matrix(x)%*%t(matrix(x))}
        res=matrix(0,b,b)
        for(k in 1:n[i])
```

```
                res=res+tempfunc((temp$y-temp$Rijdot)[temp$sub==k])
                resv=resv+res/(n[i]*(n[i]-1))  # Vjj'
                res2=res2+sum(res)/(n[i]*(n[i]-1))
                }
        list(msea=res2/(a*b),Vb=resv/(a^2))
        }


### samfun to got the value of sigma<ijj1>

samfun=function(data,a,b, coln=4){
                sgmijj1=numeric()
                for(i in 1:a){
                        for(j in 1:b){
                                for(j1 in 1:b){
                                        x=data[(data[,1]==i & data[,2]==j),coln]
                                        y=data[(data[,1]==i & data[,2]==j1),coln]
                                        sigijj1= cov(x,y)
                                        sgmijj1=rbind(sgmijj1,c(i,j,j1,sigijj1))
                                        }
                                }
                        }
                sgmijj1
                        }

###taufun is going to cal the value of tauA=(1/ab^2)*(sum(2/n(n-
1))sum(sigma<ijj1>sigma<ij2j3>

taufun=function(Data,sigma, n,a,b){

        #b = nrow(Data)
```

```
#a = length(n)
X=Data

VQ = 0
V1 = 0 # variance matrix V1 -- sum(\sigma^2_{i,jj1)) for any j, j1
V2 = 0
V3 = 0

for (i in 1:a) {
if (i==1) start = 1 else {
start = sum(n[1:(i-1)])+1
          }
end = sum(n[1:i])
temp = X[,start:end]
temp.1 = cbind(temp[,-1], temp[,1])
Xd = temp-temp.1 # paired difference X_{ijk}-X_{ijk+1}
Xd.mult1 = kronecker(Xd, rep(1,b))
Xd.mult2 = kronecker(rep(1,b), Xd)
Xd.prod1 = Xd.mult1 * Xd.mult2
Xd.mult3 = kronecker(rep(1,b^2), Xd.prod1)
Xd.mult4 = kronecker(cbind(Xd.prod1[,-c(1,2)], Xd.prod1[,c(1,2)]), rep(1, b^2))
V.prod = Xd.mult3 * Xd.mult4
VQ = VQ + sum(V.prod)/(2*n[i]^2*(n[i]-1))


V1.id = c(TRUE, rep(c(rep(FALSE, b^2), TRUE), b^2-1))
V1 = V1 + sum(V.prod[V1.id,])/(4*n[i]^2*(n[i]-1))
V2 = V2 + sum(V.prod)/(4*b^2*n[i]^2*(n[i]-1))
V3.id = c(rep(c(rep(TRUE, b), rep(FALSE, b^2-b)), b-1), rep(TRUE, b),
  rep(c(rep(FALSE, b), rep(c(rep(FALSE, b^2-b),
        rep(TRUE, b)), b)), b-1))
```

```
        V3 = V3 + sum(V.prod[V3.id,])/(2*b*n[i]^2*(n[i]-1))
        }
        tauA = VQ /(a*b^2)
        tauAB = 2* (V1 + V2 - V3) /(a*(b-1)^2)


        sgma=sum(tapply(sigma[,4], sigma[,1],sum)/n)/(a*b)
        sgmijsq=sigma[sigma[,2]==sigma[,3],]
        sgmAB1=mean(tapply(sgmijsq[,4],sgmijsq[,1],sum)/n)/(b-1)
        sgmAB2=sgma/(b-1)
        sigmaAB=sgmAB1-sgmAB2

    list(tauA=tauA, tauAB=tauAB, sigmaA=sgma, sigmaAB=sigmaAB,
resulttaufun=c(tauA=tauA, tauAB=tauAB, sigmaA=sgma, sigmaAB=sigmaAB))
    }

    ### compute the main effect and interaction
    #note:
    #       First coln of data is named trt, second coln is named time
    #       Third coln is named sub, and the last coln is y,

    test=function(Data, data, a=ncal(data)$a, b=ncal(data)$b, coln=4, n=ncal(data)$n,
contrast=cbind(diag(rep(1,b-1)),rep(-1,b-1))){

        trt=data[,1]
        time=data[,2]
        sub=data[,3]
        y=data[,coln]
        Rij=as.matrix(tapply(y, list(trt, time), mean) )
        Ri=as.matrix(apply(Rij, 1, mean) )
        Rijdot=dot(Rij,n)
```

```
        newdat=as.data.frame(cbind(data,Rijdot))
        colnames(newdat)=c(colnames(data),"Rijdot")


### main test function;
### test effect A(trt)

        mseA=msecal(newdat,a,n,b)$msea
        mstA=b*sum((Ri-mean(Ri))^2)/(a-1)
        Fa=mstA/mseA

        taufuncal=taufun(Data,sigma=samfun(data,a,b,coln),n,a,b)
        tauA=taufuncal$tauA
        sigmaA=taufuncal$sigmaA
        Va=sqrt(tauA)/sigmaA

        test.statA=sqrt(a)*(mstA-mseA)/sqrt(tauA)
        test.statA

        pA=2*(1-pnorm(abs(test.statA)))

### test effect AB (trt*time)

        Rj=as.matrix(apply(Rij, 2, mean) )

        Rdot=mean(Ri)          #Rdot is X...(tilde)
        Rjmatrix=matrix(rep(Rj,a),a,b, byrow=T)
        mstAB=sum((Rij-c(Ri)-Rjmatrix+Rdot)^2)/((a-1)*(b-1))

        mseAB1=mean(apply(tapply(newdat$y-
                newdat$Rijdot,list(newdat$trt,newdat$Time),
```

```
                function(x) sum(x^2)),1,sum)/(n*(n-1)))/(b-1)
        mseAB2=(msecal(newdat,a,n,b)$msea)/(b-1)
        mseAB=mseAB1-mseAB2

        Fab=mstAB/mseAB
        tauAB=taufuncal$tauAB
        sigmaAB=taufuncal$sigmaAB

        Vab=sqrt(tauAB)/sigmaAB

        test.statAB=sqrt(a)*(mstAB-mseAB)/sqrt(tauAB)
        test.statAB

        pvalueAB=2*(1-pnorm(abs(test.statAB)))

    ### test effect B (time)

        Vb=msecal(newdat,a,n,b)$Vb
        Wb=matrix(Rj,nr=1)%*%t(contrast)%*%solve(contrast%*%Vb%*%t(contrast))
%*%contrast%*%matrix(Rj,nc=1) #Chisq dist.
        pB=1-pchisq(Wb,nrow(contrast))

### results
    list(Fa=Fa,Fab=Fab,mstA=mstA,mseA=mseA, mstAB=mstAB,mseAB1=mseAB1,
mseAB=mseAB, test.stat.trt=test.statA, test.stat.inter=test.statAB, p.trt=pA,
p.time=pB,p.inter=pvalueAB, tauA=tauA, tauAB=tauAB, sigmaA=sigmaA,
sigmaAB=sigmaAB,  resultmseA=c(mseA=mseA), resultmseAB=c(mseAB=mseAB),
     resultsgmA=c(sigmaA=sigmaA), resultsgmAB=c(sigmaAB=sigmaAB),
resultFA=c(Fa=Fa), resultFAB=c(Fab=Fab), resultstatA=c(test.stat.trt=test.statA),
     resultstatAB=c(test.stat.inter=test.statAB),
resultPs=c(pA=pA,pB=pB,pvalueAB=pvalueAB), resultPA=c(pA=pA),
```

```
resultPB=c(pB=pB), resultVA=c(Va=Va), resultVAB=c(Vab=Vab),
resultPAB=c(pvalueAB=pvalueAB),
resultTaus=c(tauA=tauA,tauAB=tauAB))
        }
```

A4: R code for calculation of Lawley-Hotelling, Wilks' Lambda, and Bartlett-Nanda-Pillai.

```
### test treatment using Lawley-Hotelling, Wilks' Lambda, Bartlett-Nanda-Pillai

mv.trt.tests=function(dat1){
            y=dat1[,-(1:2)]
            trt=as.factor(dat1[,2])
    fit1 = manova(y ~ trt)
            old.Wilks=summary(fit1, test='Wilks')$stats[1,6]
            old.Pillai=summary(fit1, test='Pillai')$stats[1,6]
            old.Hotelling.Lawley=summary(fit1, test='Hotelling-Lawley')$stats[1,6]
            old=c( old.Pillai, old.Hotelling.Lawley, old.Wilks )

########### test from Bathke  et al.

##Bathke's H-L
a=length(unique(trt) )
ntime=ncol(y)
N=nrow(y)
B = ((N-ntime-2)*(N-a-1))/((N-a-ntime)*(N-a-ntime-3) )
K = ntime*(a -1)
D = 4 + (ntime*(a- 1) + 2)/(B-1)

#stat.Hotelling.Lawley=summary(fit1, test='Hotelling-Lawley')$stats[1,2]
```

```
F.Hotelling.Lawley=summary(fit1, test='Hotelling-Lawley')$stats[1,3]
p.Hotelling=pf(F.Hotelling.Lawley, K, D, lower.tail = FALSE)

##Bathke's Pillai
gamma=min(a-1,ntime)
v1=((ntime*(a-1))/(gamma*(N-1)))*(((gamma*(N-a+gamma-ntime)*(N+2)*(N-1))/((N-
a)*(N-ntime))-2))
v2=((N-a+gamma-ntime)/N)*((gamma*(N-a+gamma-ntime)*(N+2)*(N-1))/((N-a)*(N-
ntime))-2)
stat.Pillai=summary(fit1, test='Pillai')$stats[1,2]
F.Pillai=((stat.Pillai/gamma)/v1)/((1-stat.Pillai/gamma)/v2)
p.Pillai=pf(F.Pillai, v1, v2, lower.tail = FALSE)
old=c(old, p.Pillai, p.Hotelling)
}
```

A5: R code for calculation of LME and GEE

```
library(nlme)
calcStat.LME = function(sim.data, n) {
a = length(n)
b = nrow(sim.data)
Time =as.vector(row(sim.data))
Trt = as.vector(t(matrix(rep(rep(1:a, n), b), ncol=b)))
Sub = as.vector(col(sim.data))
CN = as.vector(sim.data)
X = cbind(Trt, Time, Sub, CN)
X = data.matrix(X)
gls.o=gls(CN~Trt+Time+Trt*Time, data=data.frame(X),corr=corSymm(form=~1|Sub))
summary(gls.o)
nlme.trt=anova(gls.o, type="marginal")$"p-value"[2]
nlme.time=anova(gls.o, type="marginal")$"p-value"[3]
```

```
        nlme.int=anova(gls.o, type="marginal")$"p-value"[4]

        nlmePvalue=cbind(nlme.trt, nlme.time, nlme.int)
        }

## GEE for the probe and time interaction.
        library(geepack)
        calcStat.GEE = function(sim.data, n) {
        a = length(n)
        b = nrow(sim.data)
        Time = as.vector(row(sim.data))
        Trt = as.vector(t(matrix(rep(rep(1:a, n), b), ncol=b)))
        Sub = as.vector(col(sim.data))
        CN = as.vector(sim.data)
        X = cbind(Trt, Time, Sub, CN)
        #X = data.matrix(X)
        family = "gaussian" #"poisson"
        gee.o=try(geese(CN~Trt+Time+Trt*Time, id=Sub,data=data.frame(X),
family=family),T)
        geePvalue.trt=summary(gee.o)$mean[2,4]
        geePvalue.time=summary(gee.o)$mean[3,4]
        geePvalue.int=summary(gee.o)$mean[4,4]

        geePvalue=cbind(geePvalue.trt,geePvalue.time,geePvalue.int)
        }
```